(1)

Technical Report 912

Investigation of Psychomotor and Spatial Abilities in Simulated Air Defense Engagements

Ilene F. Gast and David M. Johnson
U.S. Army Research Institute

September 1990

United States Army Research Institute
for the Behavioral and Social Sciences

Approved for public release; distribution is unlimited.

U.S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

JON W. BLADES
COL, IN
Commanding

Technical review by

Henry Busciglio
Jon J. Fallesen

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS --
2a. SECURITY CLASSIFICATION AUTHORITY -- 2b. DECLASSIFICATION / DOWNGRADING SCHEDULE --	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S) ARI Technical Report 912	5. MONITORING ORGANIZATION REPORT NUMBER(S) --
6a. NAME OF PERFORMING ORGANIZATION U.S. Army Research Institute 6b. OFFICE SYMBOL (If applicable) PERI-SB	7a. NAME OF MONITORING ORGANIZATION --
6c. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600	7b. ADDRESS (City, State, and ZIP Code) --
8a. NAME OF FUNDING / SPONSORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences 8b. OFFICE SYMBOL (If applicable) PERI-R	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER --
8c. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 63007A PROJECT NO. 792 TASK NO. (232) 2202 WORK UNIT ACCESSION NO. H1

11. TITLE (Include Security Classification)
Investigation of Psychomotor and Spatial Abilities in Simulated Air Defense Engagements

12. PERSONAL AUTHOR(S)
Gast, Ilene F.; and Johnson, David M.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 88/01 TO 90/09	1990, September	

16. SUPPLEMENTARY NOTATION
--

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Air defense; Engagement simulation; SHORAD; Personnel selection; Test validation; Psychomotor abilities; Spatial abilities (Continued)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This report documents research performed to evaluate the feasibility of using spatial and psychomotor tests developed for Project A for predicting engagement of aerial targets by two Air Defense Military Occupational Specialties. Data were collected from 26 16P (Chaparral) personnel and 75 16S (Stinger) personnel in their last week of advanced individual training at Fort Bliss. Predictor tests were administered nearly concurrently with criterion data collection in the Realistic Air Defense Engagement System (RADES). No relationship was found between spatial or psychomotor abilities and engagement performance. Possible explanations were discussed for the negative results, including limitations in sample size and limitations in experimental control necessitated by the need to provide a realistic simulation environment.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL David M. Johnson	22b. TELEPHONE (Include Area Code) (915) 568-4491 22c. OFFICE SYMBOL PERI-SB

DD Form 1473, JUN 86 — Previous editions are obsolete.

ARI Technical Report 912

18. SUBJECT TERMS (Continued)

Criterion measures
Predictor measures

Accession For		
NTIS GRA&I		☑
DTIC TAB		☐
Unannounced		☐
Justification		
By		
Distribution/		
Availability Codes		
Dist	Avail and/or Special	
A-1		

Technical Report 912

Investigation of Psychomotor and Spatial Abilities in Simulated Air Defense Engagements

Ilene F. Gast and David M. Johnson
U.S. Army Research Institute

Selection and Classification Technical Area
Michael G. Rumsey, Chief

Manpower and Personnel Research Laboratory
Zita M. Simutis, Director

Field Unit at Fort Bliss, Texas
Michael H. Strub, Chief

Systems Research Laboratory
Robin L. Keesee, Director

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

September 1990

Army Project Number 2Q263007A792

Manpower and Personnel

FOREWORD

The research discussed in this report was the product of a coordinated effort between two U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) units. The predictor tests were the responsibility of the predictor development team of the Selection and Classification Technical Area of the Manpower and Personnel Research Laboratory. The criterion measures were the responsibility of the crew weapons performance team of the Fort Bliss Field Unit of the Systems Research Laboratory. The predictor development team determines objective predictors of successful performance in Military Occupational Specialties, with special emphasis on ARI's Project A. The crew weapons performance team measures air defense crew performance to develop doctrine, training standards, and new weapon system procurement decisions.

The research was performed to determine the extent to which psychomotor and spatial abilities correlate with engagement performance of Stinger and Chaparral trainees. This task was initiated under the direction of General Thurman, Commanding General, Training and Doctrine Command, in December of 1987. The results of this research were briefed to General Thurman and to the Air Defense Artillery School in summer 1988, and later were presented at the December 1988 meeting of the Military Testing Association. This research was to provide a tool to identify superior candidates for recruitment into the Air Defense Artillery School's excellence track for gunners, the "Top Gun" Program.

Edgar M. Johnson

EDGAR M. JOHNSON
Technical Director

(5) BLANK

INVESTIGATION OF PSYCHOMOTOR AND SPATIAL ABILITIES IN SIMULATED AIR DEFENSE ENGAGEMENTS

EXECUTIVE SUMMARY

Requirement:

To evaluate the feasibility of using spatial and psychomotor tests developed for Project A for predicting detection, identification, and engagement of aerial targets in two Air Defense Military Occupational Specialties.

Procedure:

During the spring of 1988, predictor and criterion data were collected from 26 16P personnel and 75 16S personnel during their last week of advanced individual training at Fort Bliss. Predictor tests were administered in a group session a few days before criterion testing. Criterion data were collected during a field training exercise at the Realistic Air Defense Engagement System. Troops manned actual weapon systems against subscale models of fixed and rotary wing aircraft while criterion measures were recorded for each soldier. Scores on predictors were correlated with criterion data using the Pearson Product Moment technique.

Findings:

None of the hypothesized relationships emerged. However, a number of measurement factors may have contributed to the failure to find the expected relationships. Among the measurement issues discussed are the need for larger sample size and the need for greater control of gunner tracking time.

Utilization of Findings:

This effort was part of the Skills Selection and Sustainment program at the Air Defense Artillery School (USAADASCH) at Fort Bliss. One component was an excellence track for gunners, the "Top Gun" program, which sought to identify superior gunners and provide them with additional training and development. After the results of this pilot test were briefed, USAADASCH decided not to use spatial and psychomotor tests to select superior gunners until more research had been performed.

INVESTIGATION OF PSYCHOMOTOR AND SPATIAL ABILITIES IN SIMULATED AIR DEFENSE ENGAGEMENTS

CONTENTS

LIST OF TABLES

CONTENTS (Continued)

INVESTIGATION OF PSYCHOMOTOR AND SPATIAL ABILITIES
IN SIMULATED AIR DEFENSE ENGAGEMENTS

BACKGROUND

The former Commanding General (GEN Thurman), U.S. Army Training and Doctrine Command (TRADOC), directed proponent schools to develop accelerated initial-entry training programs for all Military Occupational Specialties (MOS) in December of 1987. Schools were urged to select soldiers possessing outstanding gunnery and leadership potential for these programs. The Excellence in Armor Program developed by the U.S. Army Armor Center was to be the model for this effort.

Accordingly, the U.S. Army Air Defense Artillery School (USAADASCH) developed the Excellence in Air Defense Artillery Programs. These include a Fast Track Program for Initial Entry Training (IET) and the Skill, Selection, and Sustainment (S3) program called "Top Gun" by USAADASCH. The Top Gun Program seeks to identify superior gunners and provide them with additional training and development. These superior gunners are to be identified by aptitude tests developed by the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI), as well as by their demonstrated motivation and leadership potential in IET.

This pilot experiment was designed to evaluate the feasibility of using spatial and psychomotor tests developed for Project A: Improving the Selection and Classification of Enlisted Personnel (Eaton, Goer, Harris, & Zook, 1984; Campbell, 1988), for predicting detection, identification, and engagement of aerial targets for two Air Defense MOS. USAADASCH has been directed by TRADOC to use spatial and psychomotor tests for selection because these psychological abilities had been shown to predict gunnery skills for tankers tested at Fort Knox (Smith & Graham, 1987; Graham, 1989). If these tests were successful at Fort Bliss, they would be incorporated into the procedures for selecting soldiers into the Top Gun Program.

During a 90-day pilot test, predictor and criterion data were collected nearly concurrently. Performance data were collected during a Field Training Exercise (FTX) at the Realistic Air Defense Engagement System (RADES) facility using actual weapon systems against subscale models of fixed and rotary wing aircraft. The pilot test sample included 16P personnel who operate the Chaparral weapon system, a vehicle mounted heat seeking missile, and 16S personnel who operate the Stinger weapon system, a shoulder mounted heat seeking missile.

METHOD

Participants

Data were collected from 26 16P personnel and 75 16S personnel in their last week of Advanced Individual Training (AIT) at Fort Bliss. All had

received training in visual recognition of aircraft. The 16S trainees had passed objective qualification standards for operation of the Stinger weapon in the Moving Target Simulator. [The Moving Target Simulator (M87A1) is a facility for training Stinger gunners in engagement skills. Targets are projected onto a dome-shaped screen along with a superimposed heat source. Target tapes vary in difficulty. Stinger gunners track and engage these targets in real time under the tutelage of Stinger instructors.]

The 16P personnel had been "familiarized" in the operation of the Chaparral weapon system. Familiarization means that 16P personnel had received as much engagement training as time, targets, and instructor resources had permitted. There is no training facility dedicated to the exercise of Chaparral gunner engagement skills. Gunners are not required, therefore, to pass an objective qualification standard for engagement proficiency.

Predictors

Psychomotor (Psmtr.) skills. Two computerized psychomotor tests were included from the Project A Battery (Peterson, 1987): One-handed Tracking and Two-handed Tracking. The former test measures steadiness and precision. For this test, the subject is presented with a path consisting of horizontal and vertical lines. A target moves along this path at a constant speed. Using a joystick, the subject must attempt to keep crosshairs centered on the target at all times. Trials differ with respect to target speed, path length, and number of segments comprising the path. The Two-handed Tracking Test measures multi-limb coordination and dexterity. Two sliding knobs replace the joystick; the left hand operates the vertical control and right hand operates the horizontal control.

For both tests, scores are computed in terms of mean distance off target. A single standardized score was computed based on performance on both tests.

Spatial (Spat.) ability. Two of the Project A paper-and-pencil tests of spatial ability were included (Peterson, 1987). The Maze Test assesses the ability to scan a complex visual field and to identify patterns within that field. Soldiers confront 24 different mazes, each with four entrance points and multiple exit points. Their task is to discern the one entrance which leads to an exit.

The Orientation Test measures the ability to maintain one's perspective or bearing with respect to some object when it and its component parts have been rotated. Each of the 24 items consists of a rotated (not right-side-up) picture within a frame. At the bottom of that frame is a circle containing a dot. The soldier is asked to rotate the frame mentally so that the circle lies at the true base of the picture (i.e., the bottom of the image had it not been rotated.) When the frame rotates, so does the dot within the circle. The soldier must indicate the position of the dot once the frame has been rotated.

Scores on these timed tests were the total number of items answered correctly.

Psychomotor/spatial composite (P/S Comp). A composite of psychomotor and spatial scores was formed by double weighting the standardized psychomotor test score and adding it to the standardized spatial ability test score. This composite, developed through regression analyses, was found in past research (e.g., Smith & Walker, 1988; Smith & Graham, 1987) to maximize predictor-criterion relationships.

Other predictors. Predictors also included the official Basic Rifle Marksmanship (BRM) scores obtained for each test participant during Basic Combat Training.

Criterion Measurement: The RADES Facility

RADES is located at Condron Field, White Sands Missile Range, New Mexico. This desert area contains mountains 10 km to the west and 60 km to the south. Visibility is usually in excess of 60 km. Skies are usually clear.

There are five weapon stations in the RADES test area, each accommodating one weapon system. Thus, RADES can test up to five separate air defense fire units simultaneously. Each weapon station consists of a crew or team and their weapon, a weapon interface, a lap-top microprocessor, and a human data collector. Also at each weapon station are loudspeakers playing battlefield sounds. These sounds serve to mask irrelevant audio cues during the engagement process. The interface collects such weapon-specific gunner actions as identification friend or foe (IFF) interrogation, acquisition, lock-on, and fire. The data collector, using the lap-top keyboard, inputs crew verbal responses such as detection, visual identification (ID), and the command to engage or cease engagement. Weapon stations are connected, via cable, to RADES control in a multi-drop communications network for the transmission of data. In addition, each data collector is connected by headset to the RADES command network. This network allows data collectors to receive specific trial-by-trial information from the Test Director at RADES Control. For the Stinger portion of this test all five weapon stations were used; for the Chaparral portion only three weapon positions were used.

Air defenders at these weapon positions engage RADES subscale rotary wing (RW) and fixed wing (FW) aircraft. The RW aircraft are non-flying one-fifth scale models of U.S. and Warsaw Pact attack and utility helicopters. The helicopters are mounted on stands, located strategically throughout the test area, and pop-up, pneumatically, under computer control from behind sand dunes at scenario-scripted distances in front of the weapon systems. The specific RW aircraft used in this experiment were the U.S. AH-1 and CH-3, and the Soviet Mi-24 and Mi-8. The specific distances employed were a full-scale equivalent of 3 km or 5 km, positioned at 11:00 o'clock, 12:00 o'clock, or 1:00 o'clock.

The RADES FW aircraft are flying one-seventh scale models of U.S. and Warsaw Pact attack aircraft. The specific FW aircraft flown in this experiment were the Soviet MiG-27, Su-17, and Su-25. These aircraft flew into the air defenders' sector of responsibility from a full-scale range of 20 km,

ingressing from either 11:00 o'clock, 12:00 o'clock, or 1:00 o'clock depending upon the scripted scenario.

The FW, radio-controlled, propeller-driven models are stored, repaired, and readied for flight at Condron Field. On each test day, they are transported into the desert out of sight of weapon system personnel, launched from a catapult launcher, and flown by expert pilots into the test range in the direction of the weapon systems. The pilots fly scenario-driven flight paths. The flight paths are displayed graphically inside the target control van and the pilots are guided by voice radio.

In RADES, FW aircraft are detected, located in x and y coordinates, tracked, and ranged using a radar system, an automated radar plotting device, and an automated software track filter running on the Flight Tracking Component (FTC) microprocessor.

RADES uses a marine radar adapted for use in this system by Science Applications International Corporation and Radar Devices, Inc. RADES flying aircraft carry radar transponders which return exactly that signal the radar is looking for. When the radar receives a transponder-returned signal, it outputs it to the automated radar plotting device, which in turn outputs range, in device-specific units of measurement, and target azimuth to the FTC computer.

The FTC computer converts the range value to RADES coordinates, filters out spurious return signals, and calculates the aircraft's position in RADES coordinates. When the aircraft's position has been calculated, the FTC computer forms a target position and status message and outputs it to the control station microprocessor. The control station microprocessor, in turn, broadcasts the message to the respective weapon station lap-tops and displays the current track position update on the control station graphics screen. Flying target position updates, transmitted from the control station computer to the weapon station lap-tops, are used by weapon station software to calculate the local weapon-to-target range at which weapon and crew events occur. This system provides a weapon-to-target range which is accurate to within plus or minus 22 meters.

Weapon stations are accurate in their time of event data capture to the nearest 250 milliseconds. This is contrasted with the track updates which occur every 2 seconds. To fill in the missing track positions and to assign correct target position at time of event occurrence, the weapon station software linearly interpolates intermediate track positions. Thus, an accurate estimate of target position at time of event occurrence for RADES training and test purposes is achieved.

Graphic and tabular feedback on the times and ranges of events, on the correctness of aircraft identification, and the effects of engagement can be returned immediately upon the end of each scenario at each weapon station. At the control station, hard copy outputs of these graphics and tabular feedback displays can be printed for any or all RADES weapon stations currently being exercised.

Criterion Measures Obtained from RADES Simulations

Gunnery performance of two-man teams was tested. Each team consisted of a Leader and a Gunner. The Leader was responsible for detecting the aircraft, identifying the aircraft as either friendly or hostile, and issuing a command to engage or cease engagement. The Stinger team Leader issued his commands to the Gunner directly, since he was standing next to him at the weapon position. The Chaparral Leader issued his commands to the Gunner via the field telephone, since the Gunner was positioned inside the closed turret of his weapon five to ten meters away. The Gunner was responsible for interrogating the aircraft with IFF, acquiring the aircraft with the weapon, tracking the aircraft, ranging the aircraft, locking onto the aircraft (Stinger only), superelevating and leading (Stinger only), and firing. Criterion measures employed by the present investigation are listed in Table 1 and described in more detail below.

Leader tasks. The following criterion measures were collected for Leaders on Chaparral and Stinger teams during engagements involving FW aircraft: (a) Range at Detection--Range of aircraft at detection response in full scale kilometers (full scale distance = measured distance times scaling factor); (b) Range at Identification--Range of aircraft at identification response in full scale kilometers; (c) Time to Identification--Time interval in seconds between detection response and identification response; and (d) Percent Identified Correctly--Number of correct identification judgments ("hostile" or "friendly") divided by number of identification judgments made.

For engagements involving RW aircraft, measures included: (a) Raw Detection Time--The time in seconds from software command to raise helicopter on stand until Leader makes detection response; (b) Time to Identification (described above); and (c) Percent Identified Correctly (described above).

Gunner tasks. The following criterion measures were collected for Gunners on Chaparral teams during engagements involving FW aircraft: (a) Tracking measures which included (1) Number of Acquisitions or the number of intermittent acquisitions of aircraft by the weapon system (acquisition of the heat source on the aircraft by the heat seeker on the missile); (2) Total Acquisition Time--Total time in seconds that weapon acquisition of aircraft is maintained; and (3) Percentage of Available Time on Target--Percent of total possible acquisition time window that acquisition is actually maintained; (b) IFF Interrogation--Did Gunner interrogate the aircraft with the Identification Friend or Foe subsystem (by pushing the IFF button)? (c) Time from First Acquisition to Fire--Time in seconds from the first acquisition of the aircraft by the weapon until the fire trigger is pulled; (d) Total Engagement Time--Time in seconds of the entire engagement from detection response to fire trigger pull; and (e) Effect or Engagement Efficiency--Number of aircraft hit ("killed") divided by the number of fire events.

Stinger teams were assessed on all of the above criteria with the exception of "Time from First Acquisition to Fire". In its place, Stinger teams performed the following two tasks: (a) Time to Lock-on--Time in seconds between the first acquisition of the aircraft by the weapon heat seeker and press of uncage bar on weapon by Gunner (which locks the weapon heat seeker

Table 1

Criterion Measures Obtained from RADES: 16P Personnel

<u>Leader Tasks on FW Trials</u>

Range at Detection
Range at Identification
Time to Identification
Percent Identified Correctly

<u>Leader Tasks on RW Trials</u>

Raw Detection Time
Time to Identification
Percent Identified Correctly

<u>Gunner Tasks on FW Trials</u>

Tracking
- Number of Acquisitions
- Total Acquisition Time
- Percentage of Available Time on Target

IFF Interrogation
Time From Acquisition to Fire
Total Engagement Time
Effect

<u>Gunner Tasks on RW Trials</u>

IFF Interrogation
Time From Acquisition to Fire
Total Engagement Time
Effect

Table 1 (Cont.)

Criterion Measures Obtained from RADES: 16S Personnel

Leader Tasks on FW Trials

Range at Detection
Range at Identification
Time to Identification
Percent Identified Correctly

Leader Tasks on RW Trials

Raw Detection Time
Time to Identification
Percent Identified Correctly

Gunner Tasks on FW Trials

Tracking
- Number of Acquisitions
- Total Acquisition Time
- Percentage of Available Time on Target

IFF Interrogation
Time to Lock-on
Time to Fire
Total Engagement Time
Effect

Gunner Tasks on RW Trials
IFF Interrogation
Time to Lock-on
Time to Fire
Total Engagement Time
Effect

onto the aircraft); and (b) Time to Fire--Time interval in seconds between press of uncage bar and pull of trigger.

During engagements involving RW aircraft, tracking data were not collected for Stinger and Chaparral teams; all other listed criteria of Gunner performance were assessed. Tracking data were not collected for RW aircraft because RW aircraft are statically positioned (i.e., they do not fly).

Weapon Systems

Three Chaparral weapon systems (M48A2) were used for each day of 16P testing. Each Chaparral had a Forward Looking Infrared subsystem, firing key, gunner commo headset, MIM72C Tracking Head Trainer with arming plug. and IFF Subsystem Training Set. Additional required equipment included three TA312 field telephones with three headsets and three spools of commo wire, and three pairs of 7 x 50 binoculars. Finally, one 24N Chaparral systems mechanic with 24N tool kit was always present during 16P testing.

For each day of 16S testing five M134 Stinger Tracking Head Trainers (THT) were used. Each Stinger THT came with five THT batteries and an IFF simulator. Additional required equipment included five armor vests and five pairs of 7 x 50 binoculars.

Procedure

Administration of predictors. This research utilized a concurrent validity design. Predictor testing took place during the last week of Advanced Individual Training (AIT) just prior to criterion measurement. Four Project A tests, consisting of two paper-and-pencil and two computerized measures, were administered in a classroom setting during 4-hour sessions to as many as 60 participants at a time.

Criterion data collection. The 16P and 16S AIT personnel were tested in RADES during their unit Field Training Exercise. This requirement meant that each data collection event took place during a weekend. Typically, weapons were set-up and calibrated on Friday. Testing took place all day Saturday, all day Sunday, and during the morning of Monday.

AIT personnel were brought to the RADES site at Condron Field by their instructors. Upon arrival at RADES the trainees were instructed as to the nature of the testing and what was specifically required of them (e.g., allocation of Leader tasks, allocation of Gunner tasks). Trainees were then assigned to groups for morning or afternoon performance testing. Within each group, trainees were assigned to weapon stations and to two-man teams. Trainees also received their weapon position assignments (i.e, Leader or Gunner).

[Typically, a new AIT graduate would not be a Leader. For purposes of this experiment, however, Leader-Gunner teams were a requirement. So trainees were assigned as "Acting" Team Leaders. This did not prove to be a problem, procedurally, since the trainees were knowledgeable and eager to perform as Leaders.]

While trainees in the morning group were having their performance as air defenders tested in RADES, those in the afternoon group were having their visual capabilities measured in the RADES visual laboratory, which was temporarily located at Condron Field. In the afternoon this procedure was reversed. (The relationship between trainee visual capabilities and performance in RADES will be discussed in a subsequent report.)

Once at a weapon position the data collector reviewed the engagement actions with the team and showed them their sector of responsibility and primary target line. Each team was responsible for defending the same 90 degree sector. All weapon positions were visually and aurally independent of one another and no cross cuing was permitted.

Each team was instructed in the discrete trial procedure employed, and reminded of the trial-begin and trial-end signals. Each trial began when the data collector gave the team an alerting and cuing message. This message stated that air activity was imminent, reminded the team that their Weapons Control Status was "tight", gave a clock azimuth cue to the predicted aircraft ingress avenue, and identified the intruder as either high or low in elevation. A typical alert was "red, tight, one o'clock, low!". (The primary target line was always twelve o'clock.) The purpose for having the Weapons Control Status set at "tight" was to force the leader to make his identification response based upon visual criteria. The data collector signalled the end of a trial by alerting the team that the current air attack had subsided ("return to status yellow").

Each team received a total of 14 data trials, one trial each on 14 different scenarios. The scenarios were grouped into two sets (A and B) of seven, each set containing three hostile FW scenarios, two hostile RW scenarios, and two friendly RW scenarios. (See Table 2.) Each team member completed both sets, performing one set as Gunner and the other as Leader. The order of trials within a set was counterbalanced across days of the experiment. One practice trial was given before each set. The aircraft used for a practice trial never reappeared during data trials. Participants received feedback on their performance after they had finished both sets.

Hypotheses

It was hypothesized that the psychomotor test scores would predict the tracking performance of the Gunner for fixed wing engagements and the spatial test scores would predict rotary wing detection times of the Leader. It was not anticipated that the spatial and psychomotor test scores would directly predict engagement efficiency (percent of aircraft detected, correctly identified, engaged, and engaged aircraft destroyed) since these predictors do not assess the requisite skills for these activities. However, better tracking performance could result in higher engagement efficiency.

Analyses

Sample descriptive statistics (N, mean, and standard deviation) of the predictor scores and engagement performance measures were computed separately for the two MOS. Also computed in each of the samples were Pearson Product

Table 2

Test Scenarios Used in this Experiment

Scen.	Type	Intent	Target	Azimuth	Aspect	Range	Duration
Set A							
1	FW	H	MiG-27	12:00	0 deg		
2	FW	H	Su-17	1:00	45 deg		
3	FW	H	Su-25	11:00	45 deg		
4	RW	F	AH-1	11:00	0 deg	3 km	25 sec
5	RW	F	CH-3	1:00	45 deg	5 km	25 sec
6	RW	H	Mi-24	12:00	45 deg	3 km	25 sec
7	RW	H	Mi-8	11:00	0 deg	5 km	25 sec
Set B							
11	FW	H	MiG-27	1:00	45 deg		
12	FW	H	Su-17	11:00	45 deg		
13	FW	H	Su-25	12:00	0 deg		
14	RW	F	AH-1	11:00	0 deg	3 km	25 sec
15	RW	F	CH-3	1:00	45 deg	3 km	25 sec
16	RW	H	Mi-24	12:00	0 deg	5 km	25 sec
17	RW	H	Mi-8	11:00	45 deg	5 km	25 sec

Moment correlation coefficients between predictor scores and performance measures.

RESULTS AND DISCUSSION

Means and standard deviations for predictors are shown in Table 3. Table 4 displays means and standard deviations for criterion scores as well as correlations between predictors and criteria for 16P and 16S personnel. The differences in criterion performance between the two MOS were subjected to inferential analysis. Table 5 shows the results of independent groups t tests between the criterion performance of 16P personnel and that of 16S personnel. These differences in criterion performance will be discussed first.

Criterion scores: 16P versus 16S. Stinger personnel consistently performed better than Chaparral personnel on Gunner tasks. Of the nine measurements recorded for Gunners, Stinger teams performed significantly better on seven. On the two remaining Gunner tasks (Total Engagement Time and Effect), the direction of the non-significant differences favored Stinger. Of the seven Leader tasks measured, only one showed a significant difference (RW Detection Time) and this difference favored Stinger. Chaparral personnel did not score significantly better than Stinger personnel on any of the 16 criterion measures.

Compared with Chaparral Gunners, Stinger Gunners tracked the aircraft with fewer breaks and with a greater percentage of tracking time "on target". Stinger Gunners required less time (less tracking time, less time from first acquisition to fire, and less total engagement time) to perform their engagement sequence, but did so with greater effect (greater percentage of assessed "kills"). Clearly, the 16S personnel were objectively more proficient at employing their weapon than the 16P personnel were at employing theirs.

The most likely explanation for this effect was the different criteria to which the two groups of personnel were trained. As noted in "Participants" (above), the 16S troops were trained to a qualification standard in the Moving Target Simulator. The 16P troops, by contrast, were merely "familiarized" with the operation of the Chaparral. [Stinger gunnery skills are entry level skills (Skill Level 10). Chaparral gunnery skills are Skill Level 20. Only Skill Level 10 tasks are taught in AIT. 16P personnel complete additional training once assigned to a line unit.]

Teams from the two MOS did not differ reliably in their performance of Leader tasks. This is as it should be. In this experiment, the tasks of detection and identification were performed by the Leader independent of the Gunner and the weapon system. All members of both MOS had received similar training in visual search strategies and were qualified in visual recognition of aircraft prior to their participation in this exercise.

Table 3

Means and Standard Deviations for Predictors for 16P and 16S Personnel

	N	Mean	SD
<u>16P Personnel</u>			
BRM	19	28.78	3.98
Psychomotor Score	26	55.27	7.95
Spatial Score	26	52.18	10.07
P/S Composite	26	162.73	21.25
<u>16S Personnel</u>			
BRM	36	30.25	3.97
Psychomotor Score	75	54.55	9.40
Spatial Score	75	52.00	10.17
P/S Composite	75	161.12	25.92

Table 4

Means and Standard Deviations of Criterion Measures and their Relationships with Predictors

	Mean	SD	Correlations			
16P Personnel						
Leader Tasks			BRM[a]	PsMtr.	Spat.	P/S Comp.
Fixed Wing (Jet) Trials						
Range at Detection	9.50	3.27	.38	.10	.08	.12
Range at Identification	3.88	2.23	-.10	.08	.18	.15
Time to Identify	23.54	10.97	.52*	.07	-.12	-.01
% Identified Correctly	84.21	25.28	.03	-.02	-.34	-.19
Rotary Wing (Helicopter) Trials						
Raw Detection Time	7.03	3.02	.18	.08	-.12	.00
Time to Identify	10.37	4.03	.04	.03	-.11	-.04
% Identified Correctly	71.50	22.66	.10	.03	-.20	-.07
Gunner Tasks						
Fixed Wing (Jet) Trials						
No. Acquisitions	4.11	2.58	.08	.38	.47*	.55*
Tot. Acquisition Time	10.74	7.74	-.27	.08	.03	.07
% Available Time on Target	49.32	23.58	-.20	.13	-.37	-.10
Time from First Acquisition to Fire	21.32	9.56	.25	.01	.11	.07
Total Engagement Time	38.96	12.71	.44	.08	.29	.22
Effect	71.79	31.80	.21	-.31	.11	-.18
Rotary Wing (Helicopter) Trials						
Time from First Acquisition to Fire	13.19	4.50	.03	.00	-.03	-.02
Total Engagement Time	16.21	5.13	.07	.15	.28	.25
Effect	31.65	34.94	-.13	-.07	-.14	-.13

Note. One asterisk (*) indicates a significance level of $p < .05$; two asterisks indicate a significance level of $p < .01$.

[a]N = 16 for FW and 19 for RW Trials for BRM; N = 19 for FW and 26 for RW trials for Psychomotor, Spatial and P/S Composite.

Table 4 (cont.)

Means and Standard Deviations of Criterion Measures and their Relationships with Predictors

	Mean	SD	Correlations			
<u>16S Personnel</u>						
Leader Tasks			BRM[a]	PsMtr.	Spat.	P/S Comp.
<u>Fixed Wing (Jet) Trials</u>						
Range at Detection	9.21	2.21	.40*	-.04	.15	.03
Range at Identification	4.06	1.64	.13	-.32**	-.22	-.33**
Time to Identify	21.05	8.89	.24	.26*	.35**	.33**
% Identified Correctly	88.29	21.13	-.15	-.27*	-.10	-.24*
<u>Rotary Wing (Helicopter) Trials</u>						
Raw Detection Time	5.45	2.66	.02	-.03	-.13	-.08
Time to Identify	10.72	3.85	-.17	.07	.28*	.16
% Identified Correctly	67.29	21.04	.04	-.15	.08	-.07
Gunner Tasks						
<u>Fixed Wing (Jet) Trials</u>						
No. Acquisitions	2.40	1.77	.06	.08	.14	.11
Tot. Acquisition Time	6.54	4.48	.07	.12	.31**	.20
% Available Time on Target	81.47	21.95	.03	-.15	-.02	-.12
Time to Lock-on	9.28	6.24	.10	.14	.23	.19
Time to Fire	3.15	2.50	-.14	-.04	-.17	-.10
Total Engagement Time	28.88	8.73	-.30	.09	-.06	.04
Effect	87.27	26.09	-.03	-.26*	-.06	-.21
<u>Rotary Wing (Helicopter) Trials</u>						
Time to Lock-on	6.78	3.45	-.01	-.18	-.04	-.15
Time to Fire	4.03	2.42	-.21	.00	-.16	-.06
Total Engagement Time	15.90	5.43	-.32	-.13	-.25*	-.19
Effect	41.05	35.74	.28	.06	.16	.11

<u>Note</u>. One asterisk (*) indicates a significance level of $p < .05$; two asterisks indicate a significance level of $p < .01$.

[a]N = 35 for FW and 36 for RW Trials for BRM; N = 68 for FW and 75 for RW for Psychomotor, Spatial and P/S Composite.

Table 5

Independent Groups t Test Analysis of Criterion Measures for 16P and 16S

Leader Tasks	Mean: 16P	Mean: 16S	Results: (2-tail)
Fixed Wing Trials (N = 86)			
Range at Detect	9.50	9.21	$t = 0.45$, $p > .10$
Range at ID	3.88	4.06	$t = -0.40$, $p > .10$
Time to ID	23.54	21.05	$t = 1.02$, $p > .10$
% ID Correctly	84.21	88.29	$t = -0.71$, $p > .10$
Rotary Wing Trials (N = 100)			
Raw Det. Time	7.03	5.45	$t = 2.51$, $p < .05$
Time to ID	10.37	10.72	$t = -0.39$, $p > .10$
% ID Correctly	71.50	67.29	$t = 0.86$, $p > .10$

Gunner Tasks	Mean: 16P	Mean: 16S	Results: (2-tail)
Fixed Wing Trials (N = 86)			
No. Acquisitions	4.11	2.40	$t = 3.34$, $p < .01$
Tot. Acq. Time	10.74	6.54	$t = 3.03$, $p < .01$
Avail. Time on Trgt.	49.32	81.47	$t = -5.56$, $p < .01$
Time from First Acq. to Fire	21.32	12.43[a]	$t = 4.83$, $p < .01$
Total Eng. Time	38.96	28.88	$t = 4.00$, $p < .01$
Effect	71.79	87.27	$t = -2.18$, $p < .05$
Rotary Wing Trials (N = 100)			
Time from First Acq. to Fire	13.19	10.81[a]	$t = 2.45$, $p < .05$
Total Eng. Time	16.21	15.90	$t = 0.25$, $p > .10$
Effect	31.65	41.05	$t = -1.16$, $p > .10$

[a]For 16S personnel, "Time from First Acquisition to Fire" was computed by adding "Time to Lock-on" to "Time to Fire".

Tests of Hypotheses

Findings for 16P. None of the hypothesized relationships emerged. With few exceptions, relationships between predictors and criteria were not significantly different from zero. One exception was the significant correlation which emerged between the number of intermittent acquisitions (breaks in tracking) and spatial ability ($r = .47$, $p < .05$). A second exception was the correlation between the time taken to identify a target as friendly or hostile and BRM ($r = .52$, $p < .05$). Both correlations are opposite from the direction that we would have expected. Thus soldiers who have more psychomotor ability tend to wait longer to identify targets and lose infrared (IR) acquisition of the target (i.e., break) more frequently.

Findings for 16S. Again, none of the hypothesized relationships emerged. However, soldiers with higher marksmanship scores (BRM) tended to detect targets at greater distances ($r = .40$, $p < .05$). Further, soldiers who have higher psychomotor and spatial test scores tend to track targets for longer periods of time and wait longer to identify targets. Psychomotor ability was negatively correlated with the range at which a Leader identified the target for trials involving jets ($r = -.32$, $p < .01$). Soldiers with more spatial ability tended to wait longer to identify both jets ($r = .35$, $p < .01$) and helicopters ($r = .28$, $p < .05$).

In the Gunner role, soldiers with better spatial test scores tended to track jets for longer periods of time ($r = .31$ $p < .01$). Finally, soldiers with better spatial scores tended to take less time to complete engagements involving helicopters ($r = -.25$, $p < .05$).

Summary of Hypothesis Testing

In summary, none of the hypothesized relationships emerged. As suggested by past research, (Mikos, Casey, & Lockhart, 1980), there may in fact be no relationship between performance in the two MOS and the specific spatial and psychomotor abilities measured during the pilot test. However, a number of factors may have contributed to the failure to find the hypothesized relationships.

Sample size. First, any possible relationships may have been obscured by the small sample size. Originally samples of 40 16P and 136 16S personnel had been projected--nearly double the working sample size. Sample size was especially problematic in the Chaparral subgroup ($N = 19$ for FW; $N = 26$ for RW). With an N of 19, the 95% confidence intervals around the observed correlations were +/- .45. For the Stinger sample ($N = 68$ for FW; $N = 75$ for RW) these confidence intervals were +/- .24.

Missing data. Second, missing data may have played a role in minimizing predictor-criterion relationships in both MOS, especially for Gunner tasks in which missing data were most prevalent. Many Gunners omitted key steps and RADES can only record those engagement actions which are actually performed. Interpolating missing scores was not feasible. An individual's mean score across trials was too variable given the small number of trials of each type and the sample size did not permit a stable estimate of performance based on

sample norms. Thus, the aggregate scores were determined to be the best available estimate of performance. Scores for each dependent variable were aggregated across similar trials (either FW or RW) by taking the arithmetic mean of the engagement measures recorded.

However, the aggregate score had its own limitations. Because trials varied in difficulty, missing data introduced error variance into scores that were aggregated over trials. For example, consider three trials of increasing difficulty. Soldier A completed the first trial successfully but did not attempt the latter two, Soldier B completed the first two trials with moderate scores, and Soldier C did well on one, moderately on the second but poorly on the remaining trial. Aggregate scores reflected only completed trials. Thus, Soldier C who completed all three trials, would receive the lowest score, whereas Soldier A, who completed one trial, would have the best score. In effect, these three soldiers each completed a different criterion measure, yet their scores were entered equally into the analyses. Owing to the small number of cases, adjustments for this problem were not practical.

Team interdependence. Another concern was that in many instances, we were attempting to predict team performance from individual attributes. The RADES testbed was designed to measure team (unit) performance. The Project A predictors were designed to predict individual performance based on individual attributes. For many of the criterion variables it was impossible to separate the performance of a Leader from that of the Gunner. For example, the total time available for a Gunner to track a target depends in large measure on how early in a trial the Leader detects the target and how late the Leader issues the visual identification and engagement command.

Selection of predictors. Finally, it is conceivable that attributes other than those currently measured by this set of predictor tests may be more predictive of performance on missile based weapons systems such as Chaparral and Stinger. Examination of relationships among the RADES criterion measures suggests that when Gunners perform the requisite actions more quickly their overall performance is more effective. In this pilot investigation, we limited the tests used to the S3 Battery that had been administered at Ft. Knox and Ft. Benning. The larger Project A Computerized Predictor Battery contains measures of perceptual speed and accuracy.

Post Hoc Analyses

Past research has shown RADES to be a valid Short Range Air Defense engagement simulation (Drewfs, Barber, Johnson, & Frederickson, 1988; Johnson, Barber, & Lockhart, 1988). Similarly, there is support for the validity of the Project A spatial and psychomotor tests in predicting Gunner performance (Smith & Graham, 1987; Smith & Walker, 1988). Therefore, the authors were reluctant to accept at face value the failure to confirm the majority of the hypotheses.

Post hoc analyses were conducted to examine the possible explanations for the findings in the current investigation. These largely descriptive analyses included examining patterns of missing data, assessing reliability across trials for selected variables, and determining the effects of Leader

performance on Gunner performance. In addition, consideration was given to the tests selected from the Project A Battery for the purpose of this pilot investigation.

Missing Data

Gunner performance was key to this investigation and a sufficient number of observations of Gunner performance were missing to raise concern about consistency of measurement across individuals. This is shown in Table 6 where patterns of missing data are presented by scenario number. Note that for both Stinger and Chaparral teams, Leader variables were relatively free from missing data. Thus, the Leader data appear to be complete.

Trials involving "friendly" aircraft (either true friends or misidentified hostiles) were more frequently subject to missing data. One cause of missing data on trials involving "friendly" aircraft was the requirement for Leaders to command Gunners to cease engagement upon identification since friendly targets do not pose a threat. Although such Leader actions were tactically and doctrinally sound, they reduced the amount of data available for analysis.

Interdependence of Leader and Gunner Roles

Another concern was the extent to which Gunner performance measures were affected by Leader performance. To this end, we examined consequences of misidentification of targets and Leader latencies and their effects on total performance.

Consequences of misidentification of targets. Target identification data were examined on a trial by trial basis. Each of the 75 Stinger teams completed 2 friendly helicopter trials and 2 hostile ones, making a total of 150 of each (300 trials total). Five hostile trials produced no identification results (i.e., the Leader had not identified the helicopter by the end of the trial) making a total of 145. Of the 150 friendly targets presented and responded to, 69 (46%) were correctly identified as "friend" while 81 (54%) were misidentified as "hostile" (see Table 7). Stinger Gunners interrogated 72 of these misidentified (but friendly) targets, tracked 57, locked-onto 65, fired on 54, and hit 26. Conversely, the Stinger Leaders were less prone to misidentify hostile aircraft. Of the 145 hostile targets presented and responded to 113 (78%) were correctly identified as "hostile" while 32 (22%) were misidentified as "friend". Stinger Gunners interrogated 26 of these misidentified (but hostile) targets, tracked 15, locked-onto 12, fired on 2, and hit none. Gunners did not fire at the remaining 30 misidentified hostile targets.

Each Stinger team also completed 3 trials in which the targets were hostile jets. Of these 225 trials (75 teams times 3 jet trials per team), 14 identifications were missed for a total of 211 trials with identification responses. Leaders correctly identified 182 (86%) of these hostile targets and misidentified 29 (14%). Gunners interrogated 27 of these misidentified targets, tracked 16, locked-onto 12, fired on 7, and hit 5.

Table 6

Patterns of Missing Data

Chaparral (N = 26)

Scenario	1/11	2/12	3/13	4/14	5/15	6/16	7/17
Detect[a]	2[b]	2	1	1	1	1	0
Identify	2	2	1	1	3	1	4
Interrogate	9	2	4	7	7	7	2
Track	13	13	13	19	20	19	18
Fire	9	12	12	19	17	11	13

Stinger (N = 75)

Scenario	1/11	2/12	3/13	4/14	5/15	6/16	7/17
Detect	6	9	9	3	1	5	9
Identify	8	9	12	5	16	7	12
Interrogate	6	10	9	8	12	11	13
Track	26	24	20	32	32	29	30
Fire	22	28	19	53	42	18	36

[a]Number given is number of cases missing

[b]Detect and Identify are Leader tasks; remaining tasks are performed by Gunner.

Table 7

Consequences of Misidentification of Targets by Trial

Stinger Teams

		Helicopter Trials					Jet Trials		
Actual ID Ldr Said		Friend Friend	Friend Host.	Host. Host.	Host. Friend	Host. No ID	Host. Host.	Host. Friend	Host No ID
N		69	81	113	32	5	182	29	14
Was Target:									
Interro-	No	9	9	11	6		6	2	
gated	Yes	60	72	102	26		176	27	
Acquired	No	38	15	25	16		35	12	
	Yes	31	66	88	16		147	17	
Tracked	No	38	24	36	17		40	13	
	Yes	31	57	77	15		142	16	
Locked	No	58	16	10	20		20	17	
	Yes	11	65	103	12		162	12	
Fired On	No	66	27	18	30		20	22	
	Yes	3	54	95	2		162	7	
Hit	No	67	55	71	32		46	24	
	Yes	2	26	42	0		136	5	

Note. Stinger teams completed a total of 150 friendly helicopter trials, 150 hostile helicopter trials and 225 hostile jet trials.

Table 7 (Continued)

Consequences of Misidentification of Targets by Trial

<u>Chaparral Teams</u>

		Helicopter Trials					Jet Trials		
Actual ID Ldr Said		Friend Friend	Friend Host.	Host. Host.	Host. Friend	Host. No ID	Host. Host.	Host. Friend	Host No ID
N		28	24	40	12	0	61	14	3
Was Target:									
Interro-	No	6	8	8	1		9	3	
gated	Yes	22	16	32	11		52	11	
Acquired	No	6	7	8	1		14	5	
	Yes	22	17	32	11		47	9	
Tracked	No	21	18	27	10		28	8	
	Yes	7	6	13	2		33	6	
Fired On	No	23	13	13	11		19	11	
	Yes	5	11	27	1		42	3	
Hit	No	25	20	33	12		32	11	
	Yes	3	4	7	0		29	3	

Note. Chaparral teams completed a total of 52 friendly helicopter trials, 52 hostile helicopter trials and 78 hostile jet trials.

A similar pattern emerged for Chaparral teams. Each of the 26 Chaparral teams completed 2 friendly helicopter trials and 2 hostile ones, making 52 of each (104 trials total). Of the 52 friendly targets presented and responded to 28 (54%) were correctly identified as "friend" while 24 (46%) were misidentified as "hostile" (see Table 7). Chaparral Gunners interrogated 16 of these misidentified friends, tracked 6, fired on 11, and hit 4. Of the 52 hostile targets presented, 40 (77%) were correctly identified while 12 (23%) were misidentified as "friend". Gunners interrogated 11 of these misidentified hostiles, tracked 2, fired on 1, and hit none.

Chaparral teams completed a total of 78 hostile jet trials (26 teams times 3 trials per team), of which 3 identifications were missed. Of these 75 hostile targets presented and responded to, 61 (81%) were correctly identified while 14 (19%) were misidentified as "friend". Gunners interrogated 11, tracked 6, fired on 3, and hit 3 of these misidentified hostiles.

These data demonstrate the extent to which the Gunner depends on the Leader for information about the target. If the Leader misidentifies the target the Gunner will treat it as it is described. In this exercise, friendly targets were more likely to be misidentified, and fratricide was a more prevalent occurrence than failure to engage hostile aircraft. Previous research in RADES has shown this propensity on the part of air defenders to misidentify friendly aircraft as hostile and then, as required by doctrine, engage them (e.g., Johnson, Barber, & Lockhart, 1988; Lockhart, Johnson, & Sanders, 1987). In addition, as previously stated, friendly and hostile targets receive very different treatment per the instruction of the Leader. Gunners will cease engagement of the target as soon in the sequence of events as that command is given.

<u>Variance in Leader reaction times</u>. Table 8 displays means and standard deviations for Leader reaction time on the key tasks of target detection and identification. Note that Leaders vary in their reaction times to detect and identify targets. This variance, although within the realm of acceptable performance (Johnson, et al., 1988), suggests that each Leader has a slightly different effect on the performance of his paired Gunner.

Adjusting for variations in Leader performance is not a simple matter. Chaparral crews and Stinger teams fight as a unit. Given the standard operating procedures and rules of engagement of these units, the performance of a Leader substantially affects the performance of a Gunner. Leaders and Gunners do not perform their respective tasks neatly in parallel (independently) or in series (sequentially). For example, in a typical engagement the Leader detects the target and "hands it off" to the Gunner, then visually identifies the target, and then issues a command to engage or cease engagement. The Gunner, after being shown the target by the Leader, gets the target in his sights, then interrogates it with IFF, then tracks it, determines its range, and may even lock-onto it. Then the Gunner continues to track the target while awaiting the visual identification and engagement or cease engagement command. If a command to fire is given, the Gunner then completes the engagement. If a cease engagement command is given, the Gunner breaks off the target and then returns to the Primary Target Line.

Table 8

Means and Standard Deviations of Key Leader Tasks by Scenario

Stinger Teams

Set A (N = 37)

Scenario		1	2	3	4	5	6	7
Raw time	$\bar{X}$	97.08	72.06	90.03	3.43	11.22	3.27	5.06
to Detect	SD	15.80	6.65	20.92	2.84	4.50	3.60	6.43
Time to	$\bar{X}$	20.24	14.50	23.91	11.49	12.62	7.32	10.55
Identify	SD	12.36	8.99	13.00	5.51	6.03	3.73	6.53

Set B (N = 38)

Scenario		11	12	13	14	15	16	17
Raw time	$\bar{X}$	78.89	90.58	100.76	4.76	4.45	2.92	4.74
to Detect	SD	11.59	15.05	31.30	4.02	6.58	6.22	4.02
Time to	$\bar{X}$	19.48	21.03	28.22	10.58	10.97	10.53	11.97
Identify	SD	9.94	12.97	12.95	6.44	5.43	5.13	5.47

Chaparral Teams

Set A (N = 13)

Scenario		1	2	3	4	5	6	7
Raw time	$\bar{X}$	106.08	77.67	101.17	4.92	9.00	4.69	7.38
to Detect	SD	17.42	11.02	16.62	2.63	10.57	1.93	3.91
Time to	$\bar{X}$	23.85	18.92	20.50	12.62	10.30	8.85	9.25
Identify	SD	20.41	9.63	15.12	4.13	5.72	5.93	6.38

Set B (N = 13)

Scenario		11	12	13	14	15	16	17
Raw time	$\bar{X}$	70.08	93.31	95.77	3.54	5.15	5.00	10.54
to Detect	SD	24.30	30.87	19.65	8.36	2.67	9.18	8.36
Time to	$\bar{X}$	19.45	16.58	32.46	11.18	11.00	11.00	8.46
Identify	SD	15.83	11.49	18.81	9.20	5.29	5.64	4.70

Notice in the above description that Gunner actions follow from and are limited by Leader actions. It follows from this that measures of Gunner performance are not independent of Leader actions. In the present research, adjusting Gunner criterion scores for Leader actions would be neither practical nor reflective of actual performance.

Correlations between Leader performance and Gunner effectiveness. Table 9a shows correlations between the time it takes for the Leader to identify the target and total engagement time. As the time taken by the Leader to identify the target increases, total engagement time increases. Statistically significant correlations were obtained for six of the seven trials for Stinger teams and for five of the seven trials for Chaparral teams. This relationship between identification time and total engagement time has been reported earlier in RADES for helicopter targets (Johnson, et al., 1988).

Table 9b shows correlations between total engagement time and team effectiveness (whether the target was hit). Teams whose total engagement times were shorter were more likely to hit the helicopter targets. Remember (Table 2) that the helicopters were only available to be hit for 25 seconds. This result did not hold for the three jet trials. There are substantial differences between the nature of the jet and helicopter trials. Helicopter targets pop-up and hover within visual and weapon range for a few seconds and disappear. Team quickness is critical. Jet targets, however, fly in from a range of 20 kilometers. Even after teams have detected the ingressing jet, they must wait until it flies within weapon range before they can engage it. Quickness is not as critical. In fact, teams which are too quick may fire while the target is still out of range and thereby miss.

Reliability of Performance Across Trials

Table 10 shows the reliability across the three jet trials for Stinger Gunners on all three tracking measures. These measures are (a) number of intermittent acquisitions, i.e., breaks in tracking, (b) total time spent tracking the target, and (c) percent of total possible tracking time that the target was actually tracked. Since it is not possible to estimate inter-trial reliability when only one trial was completed, only Gunners who completed at least two of the three trials were included in the analysis. Thus data from 62 Gunners were analyzed (26 Gunners completed all 3 trials and 36 completed 2). Based on these data, it was possible to estimate the reliability given a single trial and the reliability across all three trials.

Reliabilities across the three trials range from .32 to .49. Also in Table 10 are the number of trials needed to reach reliabilities of .60 and .80 based on the Spearman-Brown Prophecy formula (Nunnally, 1978). The results of this analysis suggest that at minimum, the number of trials required to obtain an acceptable reliability of these measures would range from four to eight (see Table 10). Existing RADES data could be used to estimate the number of trials that would have to be administered in order to yield the requisite number of completed trials necessary to obtain acceptable reliability on each tracking variable.

Table 9

Leader Latencies and Team Effectiveness

a. Relationship Between Time to Target Identification and Total Engagement Time

Scenario:	1/11	2/12	3/13	4/14	5/15	6/16	7/17
Stinger Personnel							
r	.71**	.78**	.64**	.72**	.23	.57**	.65**
N	53	58	55	21	29	58	38
Chaparral Personnel							
r	.73**	.60*	.39	.91**	.47	.67**	.59*
N	17	14	13	7	9	15	12

b. Relationship Between Total Engagement Time and Team Effectiveness

Scenario:	1/11	2/12	3/13	4/14	5/15	6/16	7/17
Stinger Personnel							
r	-.21	-.03	.00	-.85**	-.44**	-.72**	-.41**
N	52	57	55	23	32	57	40
Chaparral Personnel							
r	-.16	.36	.10	-.78*	-.72*	-.92**	-.34
N	14	12	13	7	9	15	13

Note. One asterisk (*) indicates a significance level of $p < .05$; two asterisks indicate a significance level of $p < .01$.

Table 10

Reliability of Tracking Measures for Stinger Gunners

	Estimated 1-Trial Reliability	Estimated 3-Trial Reliability	Number of trials needed for $r_{xx} = .60$	$r_{xx} = .80$
Variable				
No. Acquisitions	.18	.34	7[a]	19
Total Acquisition Time	.29	.49	4	10
% Available Time on Target	.16	.32	8	20

Note. Reliabilities were computed on a sample of 62 Stinger Gunners, 26 of whom had no missing data, 36 of whom were missing one trial.

[a]Number of trials has been rounded to the nearest whole number.

Reliability could also be increased by providing a series of identical jet trials. The three jet trials varied in aircraft model, approach azimuth, aspect angle, and flight path (see Table 2). If these systematic sources of variance were removed, reliability would, no doubt, improve. However, ability to generalize findings to other aircraft might be compromised.

An alternative to increasing the number of trials would be to place additional constraints on the tracking measure. Each Gunner was permitted to track as long as he deemed necessary, as per doctrine. Therefore, the measured tracking time varied a great deal across individuals and across trials. Since the number of intermittent acquisitions (breaks in tracking) is correlated with the time that individual spends acquiring the target ($\underline{r}$ = .59, $\underline{p}$ < .05) soldiers who tracked for longer periods of time tended to receive lower accuracy scores. This variance could be reduced by measuring the number of intermittent acquisitions made by each soldier for a set period of time (e.g., the first or last ten seconds spent tracking the target).

Selection of Tests from the Project A Battery

Based on the tests of the hypotheses and the post hoc analyses, it is reasonable to conclude that the selected Project A tests are not effective predictors of Leader performance. Post hoc analyses demonstrated the soundness of the Leader data. Since Leaders' performance is not contingent upon the actions of others, these criterion measures can be considered measures of individual performance. Therefore, use of Project A predictors was justified. In addition, since few observations of Leader performance were missing, it was appropriate to aggregate across trials.

Post hoc analyses suggest that it is not possible to draw any conclusions about the effectiveness of the Project A tests as predictors of Gunner performance. Unlike Leader performance, Gunner performance, as the analyses have shown, was not assessed independently. Gunner performance depended on the ability of the Leader to identify the target correctly and to do so in a timely manner. Under different circumstances, these tests may have predicted Gunner performance. Still, interrelationships among the criterion measures suggest that other tests, particularly those measuring reaction time, may be appropriate as well.

Predicting Leader skills. The four tests selected from the Project A Battery failed to predict Leaders' performance. However, these tests were originally selected because of their ability to predict Gunner performance in other MOS (Smith & Graham, 1987). If Leaders' performance had been the key concern, and in light of the current results, other more appropriate tests could have been selected from the Project A Battery (e.g., Target Identification, Choice Reaction Time, Simple Reaction Time).

There is evidence that other measures are more effective predictors of Leader performance. As part of a separate investigation employing the current sample, measures of vision were administered. These predictors were more effective than the Project A tests in predicting Leader performance. Visual Acuity (which is reverse scored) predicted the range at which Leaders on Stinger teams detected jets ($\underline{r}$ = -.22, $\underline{N}$ = 68, $\underline{p}$ = .04), as did Movement

Sensitivity (r = -.21, N = 68, p = .04). Movement Sensitivity also predicted the time at which Leaders identified jets (r = -.25, N = 68, p = .02). Visual Acuity also predicted the percentage of helicopters identified correctly (r = -.26, N = 75, p = .01).

Predicting Gunner skills. Research by Mikos, Casey, and Lockhart (1980) suggests that the predictor battery may have been ineffective because air defenders require different skills than Gunners who must hit land-based targets. Other tests from the Project A Battery might have been better suited for measuring abilities that may be important to performance on "fire and forget" weapons such as Chaparral and Stinger. One such measure is the Cannon Shoot test which assesses ability to determine the relative distance and speed of a moving target. In addition, the correlations between team effectiveness and total engagement time (Table 9) suggest that a test of reaction time would be as appropriate for Gunners as it would be for Leaders.

Previous Research Predicting Gunner Skills

Both research efforts which report statistically significant correlations between spatial and psychomotor abilities and gunner performance employed tankers engaging ground targets presented by computer generated imagery in an M1 conduct of fire simulator (Smith & Graham, 1987; Graham, 1989). Besides obvious differences in sample MOS and type of targets, when compared to the current (Fort Bliss) study, differences exist in predictor tests and experimental methodology.

The configuration of spatial and psychomotor predictors employed by Smith and Graham (1987) differed from those used in the Fort Bliss research. Smith and Graham performed their analyses on separate spatial and psychomotor factors in contrast to the single P/S Composite used in the present investigation. The Smith and Graham "spatial" factor included the Orientation Test used in the Fort Bliss study and three additional tests: Reasoning, Object Rotation, and a Map test. Their "psychomotor abilities" factor included the two tracking tests (One-handed Tracking, Two-handed Tracking) and the Maze test employed by the present research, as well as a Target Shoot test and a Cannon Shoot test. Note that the present research included the Maze test among the Spatial tests. Graham (1989) did employ the same spatial and psychomotor predictor tests used in the current study. His spatial-psychomotor composite, however, was not constructed according to the same computational techniques as the P/S Composite score used in the current study and described earlier in this report.

In addition, large differences between the Fort Knox and the Fort Bliss investigations can be found in the Gunners' tasks. For example, consider target acquisition, clearly a critical Gunner task. In the Fort Bliss experiment, all Gunners were individually responsible for acquiring targets with the weapon after the Leader detected them. Gunners vary in their skill at performing this task. However, because previous armor research (Graham, 1986) had shown that variability in acquisition performance greatly influenced overall Gunner performance, procedures were employed to minimize the effects of acquisition in the two Fort Knox experiments. The method used by Smith and Graham (1987, pps. 11, 12) had an armor instructor "talk" the Gunner onto a

reference point before each target appeared. This assured that the target always emerged within the field of view of the Gunner's ten power sight, making the acquisition task relatively simple and constant. Graham (1989, p. 5) took this approach even farther by employing the simulation's "synthetic tank commander". Here the simulations's software automatically (and without error) acquired the target, layed the main gun, and gave the command to fire.

The point, here, is simply that substantial differences in method exist between the Fort Knox work, which did find the expected correlations, and the present experiment, which did not. Additional research will be needed to determine whether the differences in results are a function of differential skills between the subject MOS (Mikos et al., 1980) or are merely a function of differences in research techniques. The "Recommendations" which follow should be considered before any attempts are made to address this issue.

Recommendations

RADES was designed to be a Short Range Air Defense engagement simulation testbed and prototype trainer. The unit of analysis in RADES is the crew or team. In retrospect, if more time and resources had been available, measurements taken during RADES trials could have been adapted to serve test validation better. Because of the measurement issues described herein, this research cannot be considered a fair test of the Project A predictors. A number of modifications could have led to a more conclusive pilot test.

First, the role of the Leader might have been controlled. This could have been achieved by allowing a confederate to serve as the Leader or by eliminating the role entirely. Without such controls, it is impossible to assess Gunner performance accurately with tests of individual abilities. Modifying the method in this fashion reduces the tactical realism and training value of this exercise, but does permit a test of these predictors.

Second, additional controls might have been placed on Gunner performance. If in the interests of tactical realism, Gunner tracking time varied at the discretion of the Gunner, tracking performance could have been scored for a particular, constant period of time for each soldier on each trial.

Third, resources permitting, including more trials of each type would have helped. With more trials, missing data would have had less impact on scores and reliability across trials would have increased.

Finally, additional predictors could have been incorporated into the design. These tests could have been selected from the Project A Battery and from other sources.

REFERENCES

Campbell, J. P. (Ed.). (1988). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1986 fiscal year (ARI Technical Report 792). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A198 856)

Drewfs, P. R., Barber, A. V., Johnson, D. M., & Frederickson, E. W. (1988). Validation of the Realistic Air Defense Engagement System (RADES) (ARI Technical Report 789). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A190 289)

Eaton, N. K., Goer, M. H., Harris, J. H., & Zook, O. M. (1984). Improving the selection, classification, and utilization of Army enlisted personnel: Annual report, 1984 fiscal year (ARI Technical Report 660). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A178 944)

Graham, S. E. (1986). The Unit-Conduct of Fire Trainer (U-COFT) as a medium for assessing gunner proficiency: Test reliability and utility (ARI Research Report 1422). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A169 196)

Graham, S. E. (1989). Assessing the impact of psychomotor and leadership selection tests on the Excellence in Armor program (ARI Research Report 1510). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A207 593)

Johnson, D. M., Barber, A. V., & Lockhart, J. M. (1988). The effect of target background and aspect angle on performance of Stinger teams in the Realistic Air Defense Engagement System (RADES) (ARI Technical Report 822). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A207 283)

Lockhart, J. M., Johnson, D. M., & Sanders, W. R. (1987). Manpower, personnel, and training analysis of air defense crews in RADES. Proceedings of the 26th Annual U.S. Army Operations Research Symposium, pp. 221-231.

Mikos, R., Casey, R. J., & Lockhart, J. M. (1980). Validation of research on acquisition and retention of cognitive versus perceptually oriented training materials (ARI Technical Report 514). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A130 975)

Nunnally, J. C. (1978). Psychometric Theory. (2d ed.). New York: McGraw-Hill.

Peterson, N. G. (Ed.). (1987). Development and field test of the Trial Battery for Project A (ARI Technical Report 739). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A184 575)

Smith, E. P., & Graham, S. E. (1987). Validation of psychomotor and perceptual predictors of Army Officer M-1 gunnery performance (ARI Technical Report 766). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A191 333)

Smith, E. P., & Walker, M. (November, 1988). Testing psychomotor and spatial abilities. Paper presented at the 30th Annual Conference of the Military Testing Association.